

Chasseurs d'appart'

Quelle est l'émission préférée des maisons ? "Chasseurs d'appart" !

Vers de nouvelles aventures



Les Data Scientists, ces modernes alchimistes du numérique, passent des heures interminables à jongler avec des montagnes de données, cherchant **désespérément** des tendances significatives dans un océan de bruit. Ils tentent de dompter des ensembles de données hétérogènes et souvent incohérents, **luttant contre des formats disparates et des erreurs humaines**. La pression est constante, les délais sont serrés, et les attentes sont souvent démesurées. **Cette carrière fatigante ne vous convient plus, vous ne souhaitez plus consacrer votre vie à un métier aussi rude et ingrat.**



Un jour que vous regardez **votre émission préférée, chasseur d'appart'**, pour vous remonter le moral, vous avez une **illumination** ! Mais oui, **vous voulez devenir agent immobilier !** Vous avez toujours été passionné par ce domaine et plus, vous avez envie de passer à la télé. Vous décidez de vous réorienter

dans le métier d'agent immobilier.



Au quotidien, l'agent immobilier **réalise des estimations du prix des biens immobiliers pour des propriétaires vendeurs**, rédige et diffuse des annonces immobilières, organise les visites des biens, négocie les prix de vente et réalise tout le

suivi administratif jusqu'à la signature des actes de vente.

Hum, mais **que faire pour vous réorienter** dans ce métier si noble ? **Vous n'avez aucun bagage et votre seule expérience est votre adoration pour l'émission** de Sophie Ferjani et de Stéphane Plaza. Vous avez une autre de vos idées brillantes, pourquoi ne pas vous démarquer de la concurrence grâce à votre passé de Data Scientist ! Le génie que vous êtes **décide**, pour une dernière fois, **d'user de ses talents afin de développer un outil d'estimation de prix de biens immobiliers.**



Hello regression, my old friend

Avec beaucoup d'émotions et de nostalgie, vous débutez le développement de votre outil. Vous espérez qu'il vous donnera un **tremplin vers votre nouvelle aventure**.



Vous suivez **consciencieusement** les étapes suivantes :

1. Vous vous **rapprochez d'un ou plusieurs collaborateurs** pour la réalisation du projet.
2. Vous **récupérez des données** gracieusement fournies par une agence immobilière ici. **Vous avez le choix** entre deux datasets.
3. Vous **consultez vos données** et vous **créez une feuille de route** du projet. **Énumérez les tâches** puis **estimez leur durée**. Estimez les **ressources humaines et matérielles** (nombre de personnes impliquées dans le projet, les machines utilisées pour le



développement de l'outil).

4. Vous créez votre **repository github** ainsi qu'un **tableau Trello** afin de répartir, équitablement, les tâches du projet.
5. **Traitez et nettoyez vos données.** Vous devrez traiter les données manquantes (par imputation ou suppression), les dublicats, les données aberrantes et tout autre donnée dont la modification (normalisation, standardisation, ...) est **pertinente**. **Chaque choix doit être justifié et argumenté.**
6. Réalisez une **analyse exploratoire complète de vos données** à l'aide de l'outil **Power BI**.
7. **Documentez vous sur les algorithmes de régression et rédigez un descriptif du fonctionnement d'au moins 3 de ces algorithmes.** Pensez à utiliser des schémas d'illustration. Vous avez le choix de fournir votre documentation sous forme du readme de votre repository ou d'un fichier PDF.
8. **Analysez et sélectionnez les caractéristiques (features) à modéliser.** Employez différentes méthodes afin d'apporter une solution finale (Boruta, forward feature selection, ...).
9. **Entraînez au moins 3 algorithmes de prédiction du prix** des biens immobiliers puis **évaluez et comparez leur performance.**
10. Les résultats ne vous conviennent pas ? Utilisez **la méthode du grid search** afin d'identifier le modèle (et les paramètres) avec les **meilleures performances possibles.**



11. **Déployez le modèle sélectionné sur une application Flask.** Le visuel doit être un minimum travaillé.

12. **Docker est une technologie de conteneurisation qui permet d'encapsuler une application et ses dépendances dans un conteneur léger et portable.** En créant un script Docker, vous assurez que votre **application et son environnement de développement sont reproductibles** et peuvent être **exécutés** de manière cohérente **sur n'importe quel environnement**, qu'il s'agisse d'un ordinateur local, d'un serveur de production ou d'un service cloud.

Créez un script Docker permettant d'**installer les dépendances nécessaires à votre application**, **copier les fichiers de votre application dans le conteneur**, et spécifier la commande à exécuter lorsque le conteneur démarre (par exemple, lancer votre application Flask).





Compétences visées

- Analyse de données
- Apprentissage automatique
- Gestion de projet

Rendu

L'évaluation de ce projet se fera sur deux aspects :

1. Une **présentation explicative de votre travail** sous forme de diapositives. Celle-ci doit raconter la réflexion derrière votre analyse et inclure **la feuille de route du projet** ainsi que le **Trello**.
2. Un repository github public nommé **apartment-hunter**, contenant les éléments suivants :
 - a. Un à deux **notebooks Python propres et commentés** (introduction, conclusion, etc) contenant le procédé de développement de votre outil.
Pensez à répondre à la problématique.
 - b. Votre **Dashboard Power BI, contenant votre analyse exploratoire des données**, exporté sous format consultable.
 - c. Un fichier **README.md** présentant le contexte du projet, les données et leur analyse, les algorithmes utilisés et une conclusion sur votre travail.
Pensez à inclure la veille réalisée.
 - d. Le script **.py** de votre solution déployée sur Flask.
 - e. Un fichier **.dockerfile**.

Base de connaissances



- [Feuille de route : définition, outils, exemples](#)
- [Kanban : définition et fonctionnement de cette méthode agile](#)
- [Documentation Power BI](#)
- [Types de visualisations dans Power BI](#)
- [Exemples de visuels Power BI](#)
- [Pypi.org – Boruta 0.3](#)
- [Scikit-learn – Machine Learning in Python](#)
- [Scikit-learn – Regression models](#)
- [Flask – Quickstart](#)
- [Docker Docs](#)